РЕЦЕНЗІЯ

доктора технічних наук, професора Кучука Георгія Анатолійовича на дисертаційну роботу **Горносталя Олексія Андрійовича "Ансамблевий метод ідентифікації стану комп'ютерних систем"**, подану на здобуття наукового ступеня доктора філософії з галузі знань 12 – інформаційні технології за спеціальністю 123 – комп'ютерна інженерія

Ступінь актуальності теми дисертаційної роботи. Щорічно загрози в сфері інформаційних технологій, зумовлені вторгненнями різної природи, стають все більш небезпечними. Зловмисники намагаються вплинути на роботу комп'ютерних систем, удосконалюючи методи своїх атак та використовуючи вразливості у їх роботі. В таких умовах особливо важливою є задача пошуку сучасних методів ідентифікації стану комп'ютерних систем, які дозволяють в умовах зовнішнього впливу виявляти вторгнення в комп'ютерні системи.

За таких умов найперспективнішими можна вважати методи машинного навчання, а особливо – ансамблеві беггінг-класифікатори. Вони дозволяють об'єднати результати роботи окремих базових моделей, використовуючи їх сильні сторони та компенсувавши основні недоліки, а також забезпечують стійкість до перенавчання, що особливо важливо в умовах середовища, що динамічно змінюється. Таким чином, використання ансамблевих методів у завданнях ідентифікації стану комп'ютерних систем відкриває нові можливості для створення більш надійних та адаптивних систем виявлення вторгнень, а тому тема дисертаційного дослідження Горносталя Олексія Андрійовича, є актуальною, адже відповідає потребам сучасності.

Зв'язок теми дисертаційної роботи з науковими планами, програмами, фундаментальними та прикладними дослідженнями. Тема дисертаційного дослідження відповідає планам науково-дослідних робіт кафедри «Комп'ютерна інженерія та програмування». Важливість дослідження також підтверджується тим фактом, що здобувач був відповідальним виконавцем при роботі над науково-дослідною темою «Моделі і методи обробки та захисту інформації в комп'ютерних системах», де замовником виступало підприємство ТОВ «Передові цифрові рішення», м. Харків (ДР №0122U200526).

Ступінь обґрунтованості та достовірності наукових положень, висновків та рекомендацій, сформульованих у дисертаційній роботі. В рамках дисертаційного дослідження були застосовані сучасні методологічні підходи, що дозволило сформувати теоретичні припущення та експерименти для їх перевірки, а також отримати достовірні та об'єктивні результати. У процесі

формулювання наукових положень та висновків було враховано всі аспекти досліджуваної проблеми, а також залучено актуальні наукові дані та публікації.

Висновки, зроблені на основі проведеного дослідження, супроводжуються детальним аналізом, що надає їм високого рівня достовірності. Сформовані рекомендації враховують специфіку аналізованої проблеми. Такий підхід забезпечує надійність і застосовність запропонованих рішень у реальних умовах, саме тому результати дисертаційної роботи були успішно впроваджені на практиці.

Проведений аналіз приводить до висновку, що дисертаційна робота Горносталя Олексія Андрійовича є ґрунтовним дослідженням з достовірними положеннями та висновками, а також містить важливі практичні рекомендації.

Наукова новизна положень, висновків та рекомендацій, сформульованих у дисертації. Основною метою дисертаційного дослідження є підвищення якості ідентифікації стану комп'ютерних систем шляхом розробки та удосконалення методів розпізнавання аномалій та зловживань. Для цього в роботі розглядаються беггінг-ансамблі, які дозволяють поєднувати різні методи машинного навчання. В рамках роботи сформовані рекомендації щодо вдосконалення цього підходу з метою підвищення ефективності його використання в задачах ідентифікації стану комп'ютерних систем. За результатами досліджень можна сформувати такі наукові досягнення:

Отримав подальший розвиток метод ідентифікації стану комп'ютерної системи на основі дерев рішень та мета-алгоритму беггінг за рахунок вибору оптимальних гіперпараметрів налаштування класифікатора та використання процедури попередньої обробки даних, яка сфокусована на видаленні аномальних даних та зменшенні статистичної залежності між ознаками, що дозволило підвищити якість ідентифікації стану КС.

Отримав подальший розвиток ансамблевий метод ідентифікації стану комп'ютерної системи завдяки використанню багатошарового перцептрону у якості базової моделі ансамблю та вибору оптимальних гіперпараметрів налаштування класифікатора, що дозволило підвищити якість його функціонування.

Удосконалено ансамблевий метод ідентифікації стану комп'ютерної системи на основі гомогенного мета-алгоритму беггінг за рахунок розробки спеціальної процедури зменшення кількості базових класифікаторів та їх ранжування під час зваженого голосування, що дозволило зменшити час роботи ансамблю та підвищити якість класифікації стану КС.

Вперше запропоновано метод ідентифікації стану комп'ютерної системи, який відрізняється від відомих методів використанням гетерогенного мета-

алгоритму беггінг та включає трьохетапний процес підбору базових моделей класифікатора на основі технології Pasting, що дозволило підвищити ефективність ідентифікації стану КС.

Ці елементи наукові новизни дозволили підвищити ефективність процесу ідентифікації стану комп'ютерних систем та можуть використовуватися на різних етапах формування беггінг-класифікаторів.

Наукова та практична цінність одержаних результатів. Основною метою дисертаційного дослідження є підвищення якості ідентифікації стану комп'ютерних систем шляхом розробки та удосконалення методів розпізнавання аномалій та зловживань. Розв'язання основних задач роботи дозволило краще зрозуміти основні механізми виявлення вторгнень, а також дало можливість розробити нові підходи та методики для вирішення актуальних проблем у цій галузі. Практична цінність результатів дослідження полягає в можливості їх застосування в реальних умовах, що підтверджується актами впровадження. Так, отримані результати було використано при викладанні навчальних дисциплін на кафедрі «Комп'ютерна інженерія та програмування» НТУ «ХПІ», а також в комп'ютерних системах ТОВ «Передові цифрові рішення».

Основними практичним досягненнями роботи можна вважати наступні:

1. Збільшено швидкість ідентифікації стану КС шляхом використання розглянутих етапів попередньої обробки даних.

2. Покращено значення метрики *AUC-ROC* ансамблевого класифікатору на 3% при роботі на тестовому наборі даних завдяки використанню розглянутих кроків попередньої обробки даних.

3. Збільшено точність роботи ансамблевого-класифікатору за рахунок використання багатошарового перцептрону в якості базової моделі гомогенного беггінг-ансамблю.

4. Підвищено значення F_1- *Score* за рахунок розробленого програмного забезпечення, що реалізує комплексне використання ансамблевої обрізки та зваженого голосування при агрегації результатів роботи окремих базових моделей в рамах гомогенного беггінг-класифікатору.

5. Збільшено значення показника якості класифікації F_1-*Score* завдяки використанню запропонованого методу побудови гетерогенного ансамблю, який має 3 етапи відбору базових моделей для їх подальшого об'єднання в раках беггінг-класифікатору.

Наявні результати підкреслюють високу наукову та практичну значимість дисертаційного дослідження.

Повнота викладення наукових і прикладних результатів дисертації в опублікованих працях. З метою висвітлення основних результатів досліджень

здобувачем опубліковано 5 наукових статей у фахових виданнях України, серед яких одна стаття у виданні, що входить до бази Scopus. Також здобувач активно приймав участь в багатьох міжнародних конференціях та симпозіумах, де була проведена апробація наукових результатів дисертаційного дослідження. Основні результати дисертаційної роботи у публікаціях відображено достатньо повно.

Оцінка змісту дисертації, її завершеності й оформлення. За структурою та змістом дисертаційне дослідження відповідає існуючим вимогам до роботи на здобуття наукового ступеня доктора філософії. Воно складається зі 170 сторінок, серед яких анотації у двох версіях, вступ, 4 розділи, висновки, список використаних досліджень, що має 137 найменувань, та 3 додатки.

У вступі сформульовані основі характеристик дисертаційної роботи та розглянуто питання його актуальності. Розглянуті задачі та методи дослідження, а також основні елементи наукової новизни. Підкреслено практичні результати дослідження, а також виконано огляд статей та матеріалів міжнародних конференцій, на яких вони були представлені та апробовані.

Перший розділ містить огляд актуальної статистики у сфері виявлення вторгнень різного характеру у комп'ютерні системи як на рівні України, так і в інших країнах. Здобувач розглянув існуючі види загроз та основні характеристики систем, які займаються їх виявленням. Усі види методів виявлення вторгнень у комп'ютерні системи умовно поділено на сигнатурні, що сфокусовані на пошуку зовнішньої схожості процесів з загрозами, та на евристичні, які досліджують поведінкову складову загроз. В кожній з груп розглянуто основні підвиди, а також їх переваги та недоліки. Особливу увагу приділено ансамблевим методам виявлення вторгнень, які дозволяють ефективно ідентифікувати стан комп'ютерних систем в умовах збільшення їх складності. У кінці першого розділу виконано постановку науково-технічної задачі дисертаційного дослідження та сформовано основні його завдання, які мають бути виконанні в подальших розділах.

У другому розділі дисертаційної роботи досліджено особливості використання різних складових попередньої обробки даних на ефективність роботи ансамблевих класифікаторів та інших методів машинного навчання. Крім того, розглянуто вплив вибраної процедури формування вхідних послідовностей, які використовуються для навчання окремих базових моделей беггінг-ансамблю, на показники якості його роботи.

Третій розділ містить дослідження ефективності використання різних підходів, які впливають на ефективність процедури ансамблювання у беггінг-класифікаторах, а також вплив використання багатошарового перцептрону у якості базової моделі з підбором оптимальних параметрів. Розглянуто методику

ансамблевої обрізки, а також техніки зваженого голосування зі статичними та адаптивними вагами, калібрування впевненості та особливості використання мета-ознак з мета-навчанням. Виявлено високу ефективність ансамблевої обрізки та зваженого голосування з адаптивними вагами. Досліджено різні підходи їх виконання, а також перевірено вплив їх комплексного використання на показники якості ансамблевого класифікатору.

У четвертому розділі дисертаційного дослідження розглянуто перспективи використання гетерогенних ансамблів в задачах класифікації. Розроблено метод формування такого класифікатору. Він дозволяє відібрати методи машинного навчання, перевіривши їх ефективність при роботі в рамках гомогенних ансамблів та поєднати їх за допомогою беггінг-процедури Pasting. Запропонований метод дозволяє обирати між різними моделями з різними конфігураціями та налаштуваннями. При цьому один з параметрів відповідає за кількість різнорідних методів машинного навчання, що використовуються в якості базових моделей гетерогенного ансамблю. Здобувачем виконано дослідження ефективності використання запропонованої процедури з різними налаштуваннями та різними методами машинного навчання, а також продемонстровано його ефективність при розв'язанні класифікаційної задачі.

Висновки містять сукупність основних наукових та практичних досягнень дисертації за рахунок виконаних завдань які були сформульовані у першому розділі.

Список використаних джерел містить посилання на важливу в рамках дослідження інформацію та на наукові роботи закордонних та вітчизняних дослідників у сфері машинного навчання та його застосування при виявленні вторгнень у комп'ютерні системи. Додатки складаються зі списку публікації здобувача за темою дисертації, з фрагментів програм, які використовувалися для експериментальної перевірки наявних в роботі тверджень та рекомендацій, а також з актів впровадження отриманих результатів.

Зауваження до дисертаційної роботи. Загальне враження від роботи позитивне, адже вона містить комплексні теоретичні та практичні дослідження та обґрунтовані висновки з рекомендаціями, проте є ряд зауважень:

1. Другий розділ роботи містить дослідження ефективності попередньої обробки даних. При цьому порівнюється вплив окремих етапів цієї процедури на різні методи машинного навчання. Часткове очищення даних та оптимізація атрибутів дозволяють збільшити ефективність класифікації більшості моделей, проте бажано було б додатково зосередитися на тому, чи змінюється ступінь впливу в залежності від налаштувань різних моделей, наприклад, в залежності від глибини дерева.

2. У другому розділі досліджено вплив вибору процедури формування вхідних даних та оптимальних налаштувань як окремих дерев, так і всього гомогенного ансамблю. Доречно було б приділити додаткову увагу тому, як окремі комбінації параметрів впливають на результати класифікації, адже вони по-різному змінюють структуру ансамблю та його окремих базових моделей.

3. Третій розділ дисертаційної роботи містить дослідження ефективності використання багатошарового перцептрону у гомогенних беггінг-ансамблях. За результатами дослідження сформовано рекомендації щодо подальшого використання таких ансамблів. Бажано було б додати більш детальне пояснення отриманих результатів, враховуючи складність структури багатошарового перцептрону.

4. У третьому розділі розглянуто ефективність використання різних підходів до ансамблювання та їх вплив на якість класифікації. Було виявлено, що найкращі результати вдається отримати при використанні ансамблевої обрізки та зваженого голосування. При вивченні ефективності їх комплексного застосування бажано було б розширити множину показників якості класифікації та описати вплив використаних підходів (наприклад, вплив недостатньої або надлишкової ансамблевої обрізки на якість класифікації).

5. У четвертому розділі розглянуто основні переваги та недоліки гомогенних та гетерогенних ансамблів. Було б добре посднати в табличному або графічному вигляді їх основні сильні та слабкі сторони або оцінити їх за певним набором критеріїв, що покращило б сприйняття матеріалу.

6. Запропонований у четвертому розділі метод побудови гетерогенного ансамблю дозволяє відібрати методи машинного навчання для їх подальшого об'єднання. За кількість різних моделей в рамках одного пулу відповідає показник k. За результатами дослідження його збільшення приводило до зменшення ефективності запропонованого методу, проте можливо ефект був би іншим при використанні інших (більш несхожих) базових моделей. В такому випадку, можливо, більше значення параметру k дозволило б отримати різноманітніший ансамбль, який був би стійким до перенавчання та з високим рівнем узагальнюючої здатності.

Відповідність дисертації встановленим вимогам і загальні висновки. Зазначені зауваження не впливають на загальне позитивне враження від роботи, не зменшують її якості, а також наукової та практичної цінності. Вони не є визначальними і можуть бути враховані як напрямки подальших досліджень. Під час вивчення змісту та отриманих результатів дисертаційного дослідження **не було виявлено фактів порушення академічної доброчесності**. Крім того, було

встановлена, що тема роботи є актуальною та відповідає спеціальності 123 – «Комп'ютерна інженерія.

Розглянута робота "Ансамблевий метод ідентифікації стану комп'ютерних систем" є комплексним дисертаційним дослідженням та відповідає вимогам до оформлення дисертації (затверджені Наказом МОН України від 12.01.2017 № 40) та вимогам пп. 6 – 9 Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії (затверджено Постановою КМУ від 12.01.2022 р. №44 зі змінами), а її автор, Горносталь Олексій Андрійович, заслуговує на присудження наукового ступеня доктора філософії за спеціальністю 123 – комп'ютерна інженерія.

Професор кафедри комп'ютерної інженерії та програмування
Національного технічного університету
«Харківський політехнічний інститут»
доктор технічних наук, професор Георгій КУЧУК

"____" травня 2024 р.

Підпис пмор Георгій Кучук
ЗАСВІДЧУЮ:
ВЧЕНИЙ СЕКРЕТАР
НАЦІОНАЛЬНОГО-ТЕХНІЧНОГО УНІВЕРСИТЕТУ
"ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ"
"23" 05 2024 р.
ЗАЙЦЕВ Ю. І.